1.5 Independence

The previous example considers two typical *independent* events: whether it will rain or not does not depend on whether the two colleagues forgot their umbrellas or not. We therefore expect

$$P(R|A_i) = P(R)$$

with $P(R|A_i) = P(RA_i)/P(A_i)$ which equals $P(R)$ if $P(RA_i) = P(R)P(A_i)$.

Example. Roll 2 dice, $A_1 =$ "result of die 1 is 6", $A_2 =$ "result of die 2 is 6". A_1 and A_2 are independent events with $P(A_1A_2) = 1/36 = 1/6 \cdot 1/6 = P(A_1)P(A_2)$ and $P(A_2|A_1) = P(A_2) = 1/6$. *The occurrence of* A_1 *has no effect on* A_2.

Definition. Two events A and B are statistically independent if $P(A \cap B) = P(A)P(B)$.

Theorem. Suppose $0 < P(B) < 1$.
Then A, B independent $\iff$ $P(A|B) = P(A)$.

Proof. $P(AB) = P(A)P(B) \iff P(A|B) = P(AB)/P(B) = P(A)P(B)/P(B)$ $= P(A)$. □

Example (quality control). Consider SWOR, M defectives in a lot of N, a sample of size 2 and $A_i =$ "i-th selection is defective", $i = 1, 2$. The number of possible samples (ordered) is $N(N-1)$, so

$$P(A_1) = \frac{M(N-1)}{N(N-1)} = \frac{M}{N}, \quad P(A_2) = \frac{M(M-1) + (N-M)M}{N(N-1)} = \frac{M}{N},$$

but $P(A_1A_2) = \frac{M(M-1)}{N(N-1)}) \neq P(A_1)P(A_2)$.

Definition. Events $A_1, A_2, \ldots, A_n$ are **mutually independent** if for every subcollection $A_{i_1}, \ldots, A_{i_k}$

$$P(A_{i_1} \cdots A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j}).$$

Note. This must hold for *any* subcollection, not only for pairs ("pairwise independence").

Example. Flip a fair coin three times. Consider $A =$ "the number of heads is even", $B =$ "the first two flips are all heads or all tails", $C =$ "the last two flips are heads". Are A, B and C independent?
Answer: we have pairwise independence: $P(A) = 1/2 = P(B)$, $P(C) = 1/4$, $P(AB) = 1/4 = P(A)P(B)$, $P(AC) = 1/8 = P(A)P(C)$, $P(BC) = 1/8 = P(B)P(C)$. We do not have mutual independence because $P(ABC) = 0 \neq P(A)P(B)P(C)$.

Theorem. Suppose $A_1, A_2, \ldots, A_n$ are mutually independent events; $A_{i_1}, \ldots, A_{i_k}$ and $A^c_{j_1}, \ldots, A^c_{j_m}$ are two subcollections such that $\{i_1, \ldots, i_k\}$ and $\{j_1, \ldots, j_m\}$ are disjoint. Then the events $A_{i_1}, \ldots, A_{i_k}$, $A^c_{j_1}, \ldots, A^c_{j_m}$ are independent and

$$P(A_{i_1} \cdots A_{i_k} \cap A^c_{j_1} \cdots A^c_{j_m}) = \prod_{s=1}^{k} P(A_{i_s}) \prod_{t=1}^{m} (1 - P(A_{j_t})).$$

Proof. Consider $B_1, \ldots, B_m$ independent (arbitrary subset of $\{A_1, \ldots, A_n\}$). Write $C = B_1 \cdots B_{m-1}$. We have $P(B_1 \cdots B_{m-1} B_m^c) = P(CB_m^c) = P(C) - P(B_1 \cdots B_m) = \prod_{i=1}^{m-1} P(B_i) - \prod_{i=1}^{m} P(B_i) = \prod_{i=1}^{m-1} P(B_i) \times \{1 - P(B_m)\}$. Now continue iteratively [attach $B_{m+1}^c, \ldots$]. ☐

Example. Consider flipping an unfair coin with $p = P(\text{"heads on } n^{th} \text{ flip"}) \neq 1/2$. Assume the flips are independent. Only if $p = 1/2$ can we compute by counting. [Then the outcomes are equally likely and, for example, $P(\text{first H on } 2^{nd} \text{ flip"}) = 1/4$; consider HH, HT, TH, TT.]
Now $p \neq 1/2$, but we can utilize the independence assumption to compute probabilities, for example

$P(\text{"H on first flip"}) = P(\text{"H on first flip, later anything"}) = p \cdot 1 \cdot 1 \cdots,$
$P(\text{"first H on } n^{th} \text{ flip"}) = p(1-p)^{n-1}, \quad n = 1, 2, \ldots,$
$P(\text{"}k^{th} \text{ H on } n^{th} \text{ flip"}) = p^k (1-p)^{n-k} \binom{n-1}{k-1}, \quad n = k, k+1, \ldots.$

1.6 Random variables and distributions

Definition. A **random variable** (r.v.) X is a real-valued function defined on a sample space $\mathcal{S}$. If $s \in \mathcal{S}$ is the actual outcome then X takes value $X(s)$ ($X : \mathcal{S} \to \mathbb{R}, s \mapsto X(s)$).

Notation. Use X (or $Y, Z, \ldots$) for the r.v. and x (or $y, z, \ldots$) for specific possible values. The **indicator function** of a set A is

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

Example. Roll two dice; $\mathcal{S}$ consists of 36 pairs (i, j). Examples of r.v.'s:

$X = X(i,j) = i =$ result of die 1,
$Y = Y(i,j) = i + j =$ total of the two dice,
$Z = Z(i,j) = 1_{\{i>j\}}(i,j) = 1_{\{(2,1),\ldots,(6,5)\}}(i,j) =$ indicator of "die 1 shows more pips than die 2".

Example. Flip a fair coin indefinitely and write 0 for tails and 1 for heads. Then $\mathcal{S} = \{s = (s_1, s_2, \ldots) : s_i \in \{0, 1\}\}$ (sequences of 0's and 1's). Possible r.v.'s:

$X(s) = s_1 + \ldots + s_n =$ number of heads in first n flips,
$T(s) = \min\{n : s_n = 1\} =$ number of flips until the first head is observed.

Example. $\mathcal{S} = [0, \infty)$, $X(s) = s$ (lifetime).

Notation. "$X \in A$" $= \{s : X(s) \in A\}$ with $A \subset \mathbb{R}$.

Definition. A random variable X has a **discrete distribution** ("X is discrete") if there is a finite or countably infinite set A such that $P(X \in A) = 1$. In this case X has a *probability mass function* (pmf) defined by $f_X(x) = P(X = x)$ (> 0 only for $x \in A$).

Example (cont.), $Y =$ "number of flips until the first head is observed". Let $p = 1/2$ (fair coin). The pmf is

$$f_Y(y) = \left(\frac{1}{2}\right)^y 1_{\{1,2,\ldots\}}(y) = 2^{-y} 1_{\{1,2,\ldots\}}(y).$$

Alternatively write $f_Y(y) = 2^{-y}$, $y = 1, 2, \ldots$ with the interpretation $f_Y(y) = 0$ for $y \notin \{1,2,\ldots\}$. It is not sufficient to write $f_Y(y) = 2^{-y}$.